CHAPTER 6
Subfertility factors rather than assisted conception factors affect cognitive and behavioural development of 4-year-old singletons

Pamela Schendelaar
Sacha la Bastide-van Gemert
Maas Jan Heineman
Karin J. Middelburg
Jorien Seggers
Edwin R. van den Heuvel
Mijna Hadders-Algra

Submitted
ABSTRACT

Objective To study underlying causal relationships between ovarian hyperstimulation, in vitro procedures, subfertility components and child cognition and behaviour.

Design Longitudinal, prospective follow-up study.

Setting Academic Centre.

Patients Four-year-old singletons born to subfertile couples (subfertile group, n=195), including singletons born after ovarian hyperstimulation IVF (COH-IVF, n=63), modified natural cycle IVF (MNC-IVF, n=53) and natural conception (Sub-NC, n=79). Newly recruited 4-year-old singletons born to fertile parents after natural conception (reference group, n=98).

Intervention None.

Main outcome measures Primary outcome: total intelligence quotient (IQ). Behavioural outcome: total problem T-score. Causal inference search algorithms and structural equation modelling were applied as statistical tools. These are unlike traditional statistics able to unravel underlying causal mechanisms and distinguish between confounders and intermediate effects.

Results No direct or indirect underlying causal effect was found between ovarian hyperstimulation and the in vitro procedure and child cognition and behaviour. However, direct negative causal effects were found between a) severity of subfertility (time to pregnancy, TTP) and cognition and b) presence of subfertility and behaviour. These associations were confounded by maternal age at child conception and maternal educational level (GES search algorithm, penalty discount: 0.5; model fit: $X^2 = 84.7, df = 72, P = 0.145, X^2/df = 1.176$; BIC-score = -317.0).

Conclusion Our study demonstrated no direct or indirect causal effects between ovarian hyperstimulation or in vitro procedures and cognitive and behavioural outcome in 4-year-old singletons born to subfertile couples. However, suffering from subfertility, especially from more severe subfertility – which by itself is associated with higher maternal age and educational level – negatively affects child cognition and behaviour.
INTRODUCTION

Long-term follow-up of the increasing number of children born following assisted reproductive techniques (ART) is important. ART is associated with perinatal adversities, yet, ART does not seem to affect cognitive and behavioural development during the first postnatal years. However, children may grow into developmental deficits since it takes time for developmental disorders to emerge.

Various ART-related factors could potentially interfere with development, such as ovarian hyperstimulation, the in vitro procedures, the underlying subfertility and parental characteristics.

Results of long-term studies on cognitive and behavioural development in ART children vary, partly due to difficulties in distinguishing relationships between ART and underlying characteristics of subfertility, parents and child. Cognitive and behavioural outcome of ART children has been reported as similar, to worse or better than that of naturally conceived children.

In order to investigate the influence of specific factors involving assisted conception on cognitive and behavioural outcome, we composed the Groningen ART cohort. The cohort consists of three groups of singletons born to subfertile couples following conventional controlled ovarian hyperstimulation in vitro fertilization (IVF)/intracytoplasmic sperm injection (ICSI) (COH-IVF), modified natural cycle IVF/ICSI (MNC-IVF) and natural conception (Sub-NC). We previously reported no differences in neurological, cognitive and behavioural outcome between the groups up to age 4 However, at 2 and 4 years, time to pregnancy (TTP) was associated with a less optimal neurological condition. Additionally, at 2 years the cohort children had higher scores on anxious-depressed behaviour than children in a reference group with fertile parents. This suggests that rather subfertility than ART components affect neurodevelopment. At 4 years we applied a causal inference approach to evaluate factors affecting cardiometabolic outcome. It indicated direct positive effects of COH-IVF and not of MNC-IVF on systolic blood pressure percentiles and subscapular skinfold thickness, suggesting that ovarian hyperstimulation was involved in worse cardiometabolic outcome in our 4-year-old IVF singletons.

In analogy, the primary aim of the present study is to explore the underlying causal relationships between ovarian hyperstimulation, the in vitro procedure and the combination of both and cognition and behaviour in 4-year-old singletons. Secondly, we aim to explore the underlying causal relationships between several aspects of subfertility and cognition and behaviour. We addressed the contribution of a) the presence of a history of subfertility by including a reference group of 4-year-old singletons born to fertile parents, and b) the duration of subfertility in terms of TTP, as a proxy for the severity of subfertility. We have chosen cognitive outcome as our primary outcome parameter as cognition has a stronger neurobiological basis than behaviour.
MATERIAL AND METHODS

Participants
Couples who achieved a singleton pregnancy following IVF/ICSI with a term date between March 2005 and December 2006 were recruited at the Department of Reproductive Medicine of the University Medical Center Groningen (UMCG) and were invited for participation during their third trimester. This resulted in two groups of children born following controlled ovarian hyperstimulation IVF/ICSI (COH-IVF) and born following modified natural cycle IVF/ICSI (MNC-IVF). Couples treated with donated oocytes or cryopreserved embryos were excluded. A third group was formed by naturally conceived children born to subfertile couples (Sub-NC) who had tried to conceive for at least one year.

A new retrospective reference group was recruited between December 2009 and February 2012 at six local child welfare centres. Parents of 4-year-old singletons who visited these centres for routine general health care were invited to participate. Couples who had tried to achieve pregnancy for more than one year or achieved pregnancy by any form of assisted conception were excluded.

Setting
Prenatal, perinatal and demographic information was collected two weeks postterm. Information on the causes and treatment of infertility was retrieved from medical records.

The follow-up assessments were carried out by trained assessors supervised by a neurodevelopmental expert (M.H.-A.), who were blind to the mode of conception. Blinding was not possible for the reference group, as this group was recruited separately from the subfertile groups. The assessments were carried out between February 2009 and February 2012 at the UMCG.

Measurements

Cognitive development
Cognitive development was evaluated using the Kaufman Assessment Battery for Children, second edition (K-ABC-II). This standardized instrument measures cognitive and processing abilities in children aged 3 to 18 years. Outcome is expressed in a total intelligence quotient (IQ) score and four IQ scale scores. In the present paper only the total IQ score is used. Raw test scores are normalized into global scores (mean: 100, standard deviation [SD]: 15). Reliability and validity of the K-ABC-II are good. The original American norms were applied as Dutch norms are lacking.

Behavioural development
Behavioural development was evaluated using the validated Dutch version of the Child Behavior Checklist (CBCL). The CBCL is a parental questionnaire to identify emotional and behavioural problems in 1.5 to 5-year-olds, classified into problem scales, such as anxious/depressed and attention problems. The sum of all questions results in the total...
problem scale score. Raw test scores are normalized into T-scores (mean: 50, SD: 10). Higher T-scores represent more problematic behaviour. The reliability and validity of the CBCL are good.¹⁶⁷

**Statistical analysis**

Fisher’s exact tests, Mann-Whitney U-tests or Student’s t-tests were performed to investigate differences between groups using the IBM Statistical Package for the Social Sciences (SPSS), version 20. P-values < 5% were considered statistically significant.

Two separate sets of explorative analyses – each consisting of a dozen of explorations – using causal inference search algorithms were performed: one focusing on ART treatment effects (ovarian hyperstimulation; the in vitro procedure; combination of both), the other focusing on subfertility effects (presence and severity of subfertility, recorded as TTP in half-years) on developmental outcome, while taking into account causal relations of these factors and other parental and child aspects.

Causal inference search algorithms result in (classes of) causal models that are found to be compatible with the data, based on the theory of causal graphs.²¹⁹,²²⁰ A causal model consists of a model and a graph, in which the latter describes the causal relation between the variables and consists of vertices (the variables), connected by edges, which can be oriented by arrows.²¹⁹ An oriented edge represents a direct causal effect between the two connecting vertices.²²¹

We applied Conservative Peter-Clark (CPC), Greedy Equivalency Search (GES) and Conservative Fast Causal Inference (CFCI) algorithms.²¹⁷ The algorithms differ in search approach and data assumptions and use particular thresholds (GES algorithm, penalty discounts: 0.4 to 2; CPC and CFCI algorithm, alpha values: 0.05 to 0.4) of announcing a certain effect significant.²²² All search algorithms were applied while taking into account background knowledge which was logically dictated by time constraints. The resulting graphs per se represent an observationally equivalent class of causal models, i.e. other graphs from that class are equally likely to have generated the data.²²⁰ Model fit was calculated and compared between the found classes of models, by choosing one directed acyclic graph, representing such a class. We used the $\chi^2$ test, $\chi^2$/degrees of freedom (df) and the Bayesian information criterion (BIC) to compare models.²²³ Additionally, structural equation modelling was applied to a graph that represents one of the better fitted classes of models to estimate effect sizes.²²³,²²⁴

By comparing the various graphs resulting from the performed search algorithms, the most prominent and consistent direct effects could be distinguished. Additionally, effects that are consistently not found can be considered an indication of the absence of a causal effect.²²⁰ The explorative analyses were performed using the freeware program TETRAD, version 4.03.10-6²²⁵ and the sem library in R, version 2.15.0.²²⁶
Ethical approval
The Medical Ethical Commission of the UMCG approved the study design. Parents provided written informed consent for study participation of their child.

FIGURE I. Flow chart of the Groningen ART cohort.
RESULTS

Ovarian hyperstimulation and the \textit{in vitro} procedure: the three subfertile groups

**Participation and demographic characteristics**

During the prenatal period, 89 COH-IVF children, 79 MNC-IVF children and 143 Sub-NC children were eligible for participation of which respectively 68 (76%), 57 (72%) and 90 (93%) children were included in the study.\textsuperscript{35} Generally, social, obstetrical and neonatal characteristics of participants and non-participants were similar, except for a lower maternal age in non-participating Sub-NC mothers compared to participating Sub-NC mothers ($P = 0.030$).\textsuperscript{35}

At 4 years, 5 (7%) COH-IVF children, 4 (7%) MNC-IVF children and 11 (12%) Sub-NC children were lost to follow-up (total postnatal attrition rate 9.3%, Figure I). One MNC-IVF girl died at 3 weeks after birth because of a congenital heart disorder. Intra-group characteristics for participants and non-participants were similar, except for a longer TTP in participating MNC-parents compared to non-participating MNC-parents ($P = 0.021$). The demographic characteristics of the three groups are listed in Table I.

**Cognitive and behavioural development**

Two Sub-NC children had missing data on the K-ABC-II. Eventually, data of 193 children (63 COH-IVF, 53 MNC-IVF and 77 Sub-NC) were analysed. The majority of children had total IQ scores within the normal range, except for two COH-IVF children (IQ: 82 and 79) and two MNC-IVF children (IQ: 79 and 77). The mean total IQ scores of the three ART study groups are listed in Table II.

One COH-IVF, one MNC-IVF and one Sub-NC child had missing data on the CBCL. Eventually, data of 192 children (62 COH-IVF, 52 MNC-IVF and 78 Sub-NC) were analysed. The majority of children of the ART study groups had total problem T-scores within the normal range, except for two COH-IVF children and 4 Sub-NC children. The mean total problem T-scores of the three ART study groups are listed in Table II.

**Causal inference approach**

The causal graph that had the best model fit was the causal model found as a result of running the GES algorithm with a penalty discount of 0.5 ($X^2 = 98.5$, df=94, $P = 0.356$, $X^2$/df= 1.048, BIC-score= -389.1). This model indicated the absence of direct effects of COH-IVF or IVF on cognitive or behavioural outcome given the other variables in the model. A direct effect of COH-IVF or IVF was also absent in causal graphs resulting from other searches (data not provided).
### TABLE I. Characteristics of participating parents and children.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>COH - IVF</th>
<th>MNC - IVF</th>
<th>Sub - NC</th>
<th>Subfertile group</th>
<th>Reference group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 63</td>
<td>n = 53</td>
<td>n = 79</td>
<td>n = 195</td>
<td>n = 98</td>
</tr>
<tr>
<td>Child characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male gender, n (%)</td>
<td>34 (54.0)</td>
<td>26 (49.1)</td>
<td>41 (51.9)</td>
<td>101 (51.8)</td>
<td>54 (55.1)</td>
</tr>
<tr>
<td>First born, n (%)</td>
<td>43 (68.3)</td>
<td>37 (69.8)</td>
<td>49 (62.0)</td>
<td>129 (66.2)</td>
<td>46 (46.9)</td>
</tr>
<tr>
<td>Corrected age at examination at 4 years of age, median (range)</td>
<td>50 (47.5 - 60.1)</td>
<td>48.9 (48.0 - 52.5)</td>
<td>48.9 (47.9 - 56.4)</td>
<td>48.9 (47.5 - 60.1)</td>
<td>49.1 (48.0 - 54.6)</td>
</tr>
<tr>
<td>Birth characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational age (weeks), median (range)</td>
<td>39.4 (33.4 - 42.3)</td>
<td>40.1 (34.6 - 42.6)</td>
<td>40.0 (30.1 - 42.7)</td>
<td>40.0 (30.1 - 42.7)</td>
<td>40.1 (32.0 - 42.4)</td>
</tr>
<tr>
<td>Preterm birth (&lt; 37 weeks), n (%)</td>
<td>7 (11.1)</td>
<td>6 (11.3)</td>
<td>5 (6.3)</td>
<td>18 (9.2)</td>
<td>3 (3.1)</td>
</tr>
<tr>
<td>Birthweight (grams), mean (SD)</td>
<td>3393.1 (563.2)</td>
<td>3384.4 (585.7)</td>
<td>3577.9 (519.4)</td>
<td>3465.6 (557.2)</td>
<td>3599.7 (507.0)</td>
</tr>
<tr>
<td>Low birthweight (&lt; 2500 grams), n (%)</td>
<td>3 (4.8)</td>
<td>4 (7.5)</td>
<td>3 (3.8)</td>
<td>10 (5.1)</td>
<td>2 (2.1)</td>
</tr>
<tr>
<td>Small for gestational age, n (%)</td>
<td>0 (0)</td>
<td>3 (5.7)</td>
<td>1 (1.3)</td>
<td>4 (2.1)</td>
<td>4 (4.3)</td>
</tr>
<tr>
<td>Signs of fetal distress, n (%)</td>
<td>19 (30.2)</td>
<td>15 (28.3)</td>
<td>34 (43.0)</td>
<td>68 (34.9)</td>
<td>29 (29.6)</td>
</tr>
<tr>
<td>Neonatal characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apgar score 5 min &lt; 7, n (%)</td>
<td>0</td>
<td>0</td>
<td>1 (1.3)</td>
<td>1 (0.5)</td>
<td>3 (4.2)</td>
</tr>
<tr>
<td>Neonatal intensive care admission, n (%)</td>
<td>1 (1.6)</td>
<td>2 (3.8)</td>
<td>5 (6.3)</td>
<td>8 (4.1)</td>
<td>10 (10.2)</td>
</tr>
<tr>
<td>Parental characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age at conception in years, median (range)</td>
<td>32.3 (26.3 - 40.9)</td>
<td>32.8 (25.3 - 37.5)</td>
<td>33.0 (22.2 - 40.3)</td>
<td>32.8 (22.2 - 40.9)</td>
<td>30.4 (18.8 - 40.5)</td>
</tr>
<tr>
<td>Paternal age at conception in years, median (range)</td>
<td>35.4 (27.5 - 56.1)</td>
<td>34.0 (28.3 - 47.8)</td>
<td>35.0 (25.5 - 48.7)</td>
<td>35 (25.5 - 56.1)</td>
<td>32.6 (22.5 - 45.1)</td>
</tr>
<tr>
<td>Education level mother (high), n (%)</td>
<td>20 (31.7)</td>
<td>20 (37.7)</td>
<td>37 (46.8)</td>
<td>77 (39.5)</td>
<td>52 (53.1)</td>
</tr>
<tr>
<td>Education level father (high), n (%)</td>
<td>26 (46.7)</td>
<td>17 (32.7)</td>
<td>29 (38.7)</td>
<td>74 (38.7)</td>
<td>53 (56.4)</td>
</tr>
<tr>
<td>Smoking during pregnancy, n (%)</td>
<td>7 (11.1)</td>
<td>7 (13.2)</td>
<td>9 (11.4)</td>
<td>23 (11.8)</td>
<td>5 (5.1)</td>
</tr>
<tr>
<td>Alcohol consumption during pregnancy, n (%)</td>
<td>3 (4.8)</td>
<td>0</td>
<td>2 (2.5)</td>
<td>5 (2.6)</td>
<td>4 (4.1)</td>
</tr>
<tr>
<td>Divorced, n (%)</td>
<td>1 (1.6)</td>
<td>1 (1.9)</td>
<td>3 (3.8)</td>
<td>5 (2.6)</td>
<td>11 (11.3)</td>
</tr>
<tr>
<td>Fertility parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to pregnancy in years, median (range)</td>
<td>4.0 (0.0 - 13.5)</td>
<td>4.0 (0.5 - 13.9)</td>
<td>2.5 (0.5 - 11.5)</td>
<td>3.5 (0.0 - 13.5)</td>
<td>0.5 (0.5 - 1.0)</td>
</tr>
<tr>
<td>Type of infertility (primary), n (%)</td>
<td>35 (55.6)</td>
<td>32 (60.4)</td>
<td>41 (51.9)</td>
<td>108 (55.4)</td>
<td>n.a.</td>
</tr>
<tr>
<td>ICSI performed, n (%)</td>
<td>41 (65.1)</td>
<td>26 (49.1)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Mann-Whitney U-tests, Student's t-test or Fisher's exact tests were performed to investigate differences between groups. *P < 0.05; **P < 0.01; ***P < 0.001.

Note: COH-IVF: infants born following controlled ovarian hyperstimulation IVF or ICSI, MNC-IVF: infants born following modified natural cycle IVF or ICSI, and Sub-NC: naturally conceived infants born to subfertile parents. In the subfertile group COH-IVF, MNC-IVF and Sub-NC groups are taken together.

*Birth weight for gestational age is < -2 standard deviations compared with the Dutch reference population (Dutch reference tables, perinatal Registration Netherlands).

Signs of fetal distress denoted by meconium stained amniotic fluid and/or cardiotocographic signs and/or acidosis.

University education or vocational colleges.

TTP was recorded in half-years. In case of a miscarriage the onset of TTP restarted, therefore TTP may be shorter than one year.

Overall missing data: gestational age n=1, preterm n=1, birthweight n=4, low birthweight n=3, small for gestational age n=1, Apgar score 5 min < 7 n=31, maternal age at conception n=1, paternal age at conception n=8, education level father n=8, time to pregnancy n=1.
Subfertility factors rather than male assisted conception factors affect cognitive and behavioural development of 4-year-old singletons.

### Cognitive development

<table>
<thead>
<tr>
<th></th>
<th>Reference, mean (SE)</th>
<th>Reference, mean (SE)</th>
<th>Mean difference [95% CI]</th>
<th>Reference, mean (SE)</th>
<th>Reference, mean (SE)</th>
<th>Mean difference [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total IQ</strong></td>
<td>108.9 (1.22)</td>
<td>-3.7 [7.8; 0.5]</td>
<td>-28 [6.8; 1.2]</td>
<td>110.4 (1.10)</td>
<td>-3.5 [-6.3; -0.7]</td>
<td></td>
</tr>
<tr>
<td><strong>Sequential IQ</strong></td>
<td>95.9 (1.31)</td>
<td>1.2 [3.0; 5.4]</td>
<td>0.5 [3.5; 4.5]</td>
<td>99.6 (1.18)</td>
<td>-3.2 [-6.0; -0.3]</td>
<td></td>
</tr>
<tr>
<td><strong>Simultaneous IQ</strong></td>
<td>115.5 (1.59)</td>
<td>-1.5 [6.3; 3.4]</td>
<td>-2.4 [7.0; 2.2]</td>
<td>117.8 (1.48)</td>
<td>-3.5 [-6.9; -0.0]</td>
<td></td>
</tr>
<tr>
<td><strong>Learning IQ</strong></td>
<td>98.3 (1.26)</td>
<td>-2.6 [6.7; 1.5]</td>
<td>-1.5 [5.4; 2.4]</td>
<td>99.6 (1.02)</td>
<td>-2.4 [-5.2; 0.3]</td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge IQ</strong></td>
<td>112.7 (1.12)</td>
<td>-6.0 [-10.3; -1.8]</td>
<td>-4.4 [8.6; 0.4]</td>
<td>111.3 (1.26)</td>
<td>-1.7 [-4.7; 1.4]</td>
<td></td>
</tr>
</tbody>
</table>

### Behavioural development

<table>
<thead>
<tr>
<th></th>
<th>Reference, mean (SE)</th>
<th>Reference, mean (SE)</th>
<th>Mean difference [95% CI]</th>
<th>Reference, mean (SE)</th>
<th>Reference, mean (SE)</th>
<th>Mean difference [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CBCL total problem T-score</strong></td>
<td>47.8 (1.06)</td>
<td>-1.6 [-4.8; 1.5]</td>
<td>-0.5 [3.4; 2.4]</td>
<td>44.4 (0.99)</td>
<td>2.2 [-0.0; 4.5]</td>
<td></td>
</tr>
<tr>
<td><strong>CBCL internalizing T-score</strong></td>
<td>48.0 (1.14)</td>
<td>-1.3 [-4.8; 2.3]</td>
<td>-3.0 [4.3; 0.4]</td>
<td>45.0 (1.04)</td>
<td>1.8 [-0.7; 4.2]</td>
<td></td>
</tr>
<tr>
<td><strong>CBCL externalizing T-score</strong></td>
<td>48.7 (1.01)</td>
<td>-1.2 [-4.2; 1.9]</td>
<td>-0.5 [3.4; 2.4]</td>
<td>46.2 (0.95)</td>
<td>2.0 [-0.2; 4.2]</td>
<td></td>
</tr>
</tbody>
</table>

Note: COH-IVF: infants born following controlled ovarian hyperstimulation, IVF or ICSI; MNC-IVF: infants born following modified natural cycle IVF or ICSI, and Sub-NC: naturally conceived infants born to subfertile parents. In the subfertile group COH-IVF, MNC-IVF and Sub-NC groups are taken together. CBCL: Child Behavior Checklist.
Presence and severity of subfertility: inclusion of the fertile reference group

Participation and demographic characteristics

Parents of 215 eligible singleton reference children were invited to participate in the present follow-up study. Ninety-eight (46%) parents allowed their child to participate in the reference group. Demographic characteristics such as sex, gestational age, firstborn, maternal age and parental educational levels were similar for participants and non-participants of the reference group (data not provided). Demographic characteristics of participants of the subfertile and reference groups are listed in Table I.

Cognitive and behavioural development

Two children of the subfertile group and one reference child had missing data on the K-ABC-II. Eventually, data of 290 children (subfertile group: 193, reference group: 97) were included in the analyses. Four children of the subfertile group had impaired total IQ scores, whereas none of the reference children had. The mean total IQ scores of the groups are listed in Table II.

Three children of the subfertile group and two reference children had missing data on the CBCL questionnaire. Eventually, data of 288 children were included in the analyses. The majority of children in both groups had total problem T-scores within the normal range, except for six children of the subfertile group and four reference children. The mean total problem T-scores of the subfertile group and the reference group are listed in Table II.

Causal inference approach

The causal graph that presented the class of found models with the best model fit as a result of the causal effect search algorithms is represented in Figure II ($\chi^2 = 84.7$, df=72, $P = 0.145$, $\chi^2/df=1.176$, BIC-score=-317.0). It indicates that the presence of subfertility did not have a direct or indirect effect on cognition. Similar results were found in resulting causal graphs from other searches (data not provided). The causal graph suggests that cognitive outcome was directly affected by two other factors: TTP and maternal smoking during pregnancy. A longer TTP was associated directly with a lower total IQ score. This effect was confounded by high maternal education through various causal paths: one path ran from maternal education via maternal and paternal age, via TTP to IQ, whereas another path ran from maternal education via smoking during pregnancy to IQ. The other direct effect on the child’s total IQ score was the negative effect of maternal smoking during pregnancy, an effect that again was confounded by maternal education. Taking this complexity of the causal pathways into account and correcting for maternal education level, the direct, unconfounded effect of TTP on total IQ score is estimated to be -0.712 (Figure II).

The causal graph suggests that the presence of subfertility had a direct adverse effect on behavioural outcome (Figure II). As such, subfertility status acted as mediator on three different paths running from high maternal educational level on total problem T-score. These paths ran either via maternal age at child conception, or via TTP, or via a combination of maternal age and TTP, including an effect of paternal age at child conception (Figure II).
It also indicates that maternal age directly affected behavioural outcome: a higher maternal age was associated with less behavioural problems in the child. Also in this path, maternal age mediated the effect of high maternal educational level. Subsequently, maternal age at conception and high maternal educational level are considered to be confounders for the effect of subfertility status on the total problem T-score. The unconfounded, direct effect of the presence of subfertility on the total problem T-score is estimated as follows: conditionally on maternal age at conception, high maternal educational level and/or TTP, the direct effect of subfertility status is estimated to be $3.003$ for less optimal behaviour.

The causal graph also suggests a relationship between the two outcome parameters, cognition and behaviour: an increase in total IQ score with one point was associated with a decrease in total problem T-score with $0.120$ points, indicating better behaviour in case of a more optimal cognitive development (Figure II). According to the causal implications, TTP has an indirect (mediated by subfertility status and total IQ score) effect on more problematic behaviour (Figure II).

**DISCUSSION**

The present study indicates that in 4-year-old singletons born to subfertile couples direct or indirect causal effects between ovarian hyperstimulation or the in vitro procedure on the one hand and cognitive and behavioural outcome on the other hand are absent. However, a direct causal relationship was found between the severity of subfertility (TTP) and cognition at age 4. Also, a direct negative causal relationship between the presence of subfertility and behaviour was found. Both effects were confounded by maternal age at child conception and maternal educational level. TTP had a direct negative effect on the child’s cognitive outcome, whereas the presence of subfertility had a direct adverse effect on behaviour. Given the finding that cognition and behaviour are directly related to one another (with behaviour being influenced by cognition) our results suggest that the severity of subfertility has an indirect – through subfertility status as well as through IQ – effect on the child’s behavioural scores, i.e. was associated with more behavioural problems.

Our findings strengthen the notion that aspects of subfertility rather than components of ART are involved in determining cognitive and behavioural outcome of children born after IVF/ICSI. Our statistical approach also revealed important confounders for these effects, especially maternal age and education. In turn, the effect of the presence of subfertility on behavioural outcome was confounded by TTP. Our results imply that suffering from subfertility per se – especially from more severe subfertility, which in turn is affected by a higher maternal age and higher educational level – negatively affects the child’s cognitive and behavioural outcome.
The fact that already known and generally expected mechanisms were also detected (Figure II), underlines the general robustness and validity of our results. Not only the causal graph with the best model fit of the GES search algorithm, but also other graphs resulting from the GES searches with a good model fit showed similar mechanisms concerning the abovementioned effects, indicating a certain consistency of the found relations between the variables. Similar effects regarding maternal age, maternal educational level and TTP on the outcomes were also found with the CPC search algorithm (model fit indices for the CPC-result with the best model fit: alpha value 0.15; model fit: $X^2= 72.1$, df=68, $P=0.344$, $X^2/df= 1.060$ and BIC-score=307.3) The fact that similar mechanisms are revealed by two different algorithms in terms of search approach and data assumptions, underline the validity of found mechanisms. One might have expected to find direct effects of birthweight and gestational age on cognitive and behavioural outcome, however, our analyses did not result in such effects (Figure II). Birthweight is indeed associated with the child total problem T-score via maternal age at conception, but this association disappeared after correction for...
maternal age. This is in line with our expectations: our groups contain few preterm and low birthweight infants. Therefore, it seems more probable that our analyses resulted in effects of subfertility and maternal factors on developmental outcome, rather than in effects of birthweight and gestational age.

To our knowledge, this is the first study in the context of neurodevelopment in ART children using causal inference search algorithms combined with structural equation modelling. With that, this is the first study that is able to distinguish between direct and indirect causal effects and to detect and correctly adjust for confounders at the same time. Our findings contribute to the clarification of the current evidence on long-term cognitive and behavioural outcome of children born following ART; an evidence that still needs further proof. Our findings that the severity of subfertility plays a role in cognitive development is in line with Zhu et al., who reported that longer TTP may be associated with a delay in achieving certain milestones, in particular those involved in cognitive and language development. However, in another study by Zhu et al. the authors were not able to demonstrate a subfertility effect on child behaviour.

To intelligibly answer our research questions, we performed two separate sets of explorative analyses: one to zoom in on effects of ART treatment, the other to focus on the effects of presence and severity of subfertility on developmental outcome. The first exploration can only be translated to its particular (smaller) subfertile population and its findings do not extend to the different population that was used for the second exploration. In the latter analysis a reference group was added, allowing for the in concert evaluation of the effect of the presence of subfertility and that of its severity.

The explorative analyses were restricted to a certain amount of variables which we have chosen a priori based on literature. Moreover, the search algorithms are particularly suitable for continuous values and less for categorical values. We have performed several analyses with the variables parity, the presence of siblings, nursery and bilingualism taken into account. These variables did not contribute to the model and would not have altered the interpretation of the models rather than unnecessarily complicate it. The same held true for the variable ICSI. We primarily aimed to explore the underlying causal relationships between ovarian hyperstimulation and the in vitro procedure and developmental outcome and we were not primarily interested in the effect of ICSI in addition to IVF.

Besides the additional value of the statistics applied, the present study has some other important strengths. Our study design enabled us to study the effects of separate components of assisted conception such as ovarian hyperstimulation, the in vitro procedure and subfertility-related aspects on the child’s cognitive and behavioural outcome. Additional strengths of the ART cohort part of the study are the attrition rate of 9.3%, the blinding of assessors to the mode of conception and the prospective design.

It must be realized that the statistical tools used in our study are explorative in nature and especially serve as indications for new research hypotheses, meaning that our results need to be interpreted with appropriate caution. However, this caution is not restricted to search algorithms alone: applying sets of multivariable regression analyses is in fact also
explorative in nature. Due to the explorative nature of our analyses a post hoc power calculation is irrelevant. Previously group sizes of the Groningen ART cohort was based on neurological outcome at 18 months. As a general rule for a structural equation model, ten or more observations for each variable may be considered reliable, a criterion that we have met.

A more general limitation of the study is the composition of the reference group which we recruited retrospectively. Consequently, we were not blinded to group status of the reference children. Parents who are concerned about their child’s health in general are more likely to participate in developmental studies. However, this means that the associations reported in the present study may have been underestimated. The relatively low participation rate in the reference group may be an indicator for selection bias.

Another issue is that our results cannot be generalized to multiples. Being a member of a twin is associated with an increased risk for developmental problems, and ART is known to be associated with multiple births.

To conclude, our study implies that suffering from subfertility per se, and especially from more severe subfertility – which by itself is affected by higher age and high educational level of the mother – negatively affects the child’s cognitive and behavioural outcome. Long-term follow-up of development and growth of children born from subfertile couples, in particular when ART is applied, remains important, given the steadily increasing prevalence of subfertility in modern society.

ACKNOWLEDGEMENTS

We thank participating parents and children for their cooperation and enthusiasm during the assessments; Arend F. Bos, M.D. Ph.D., for his help in initiating the study and Maaike Haadsma, M.D. Ph.D., for her help in including the participants; Rosan Aapkes, M.D., Elise Bennink, M.Sc., Hanneke Geut, M.D., Marjolein Jongbloed-Pereboom, M.Sc. and Bertine de Vries, M.Sc., for support in collecting and organizing the data; Michiel Schrier, M.Sc., Linze Dijkstra B.Sc., and Loes de Weerd for technical assistance.